



## DÜZGÜN GEÇİŞ OTOREGRESİF MODELLERİN BELİRLENMESİNE YÖNELİK MELEZ KODLAMALI GENETİK ALGORİTMA\*

### HYBRID CODED GENETIC ALGORITHM TOWARDS DETERMINING SMOOTH TRANSITION AUTOREGRESSIVE MODELS\*

Serkan TAŞTAN\*\*

Nilgün ÇİL\*\*\*

#### Özet

Bu makalede, düzgün geçiş otoregresif (STAR) modellerin belirlenmesi süreci bir optimizasyon problemi olarak genetik algoritmalar ile ele alınmıştır. Bu bağlamda melez kodlamalı bir genetik algoritma kullanılmıştır. Genetik algoritma tabanlı yaklaşımın başarısı, geleneksel yöntemle belirlenen bir referans STAR modeli üzerinden değerlendirilmiştir. Sunulan yaklaşım ile referans modelden daha iyi uyuma sahip modeller elde edilmiştir.

**Anahtar Kelimeler:** Düzgün Geçiş Otoregresif, STAR, Genetik Algoritma, Model Belirleme  
**Jel Sınıflaması:** C51, C63

#### Abstract

In this paper, the specification of the smooth transition autoregressive (STAR) models as an optimization problem has been handled with genetic algorithms. In this context a hybrid-coded genetic algorithm is used. The success of the genetic algorithm based approach is evaluated via a benchmark STAR model determined by conventional method. Better-fitted models than the benchmark model are obtained with the proposed approach.

**Keywords:** Smooth Transition Autoregressive, STAR, Genetic Algorithms, Model Specification  
**Jel Classification:** C51, C63

\* Bu makale, Prof. Dr. Nilgün ÇİL danışmanlığında, Serkan TAŞTAN tarafından hazırlanan “Genetik Algoritma Temelinde Rejim Değişikliği Modellerinin Belirlenmesi” başlıklı doktora tezinden türetilmiştir.

\*\* Yrd. Doç. Dr., Cumhuriyet Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Yönetim Bilişim Sistemleri Bölümü, Sivas, stastan@cumhuriyet.edu.tr

\*\*\*Prof. Dr. , İstanbul Üniversitesi, İktisat Fakültesi, Ekonometri Bölümü, İstanbul, nilgun.cil@istanbul.edu.tr

## 1. GİRİŞ

İlk defa Chan ve Tong (1986) tarafından önerilen daha sonra Teräsvirta (1994) tarafından geliştirilen düzenli geçiş otoregresif modellerinde, doğrusal otoregresif modeller verinin dinamik özelliklerine göre tanımlanan bir sınırlandırılmış geçiş fonksiyonu ile bağlanmıştır (Chen, 2003). Sürekli bir fonksiyon olan geçiş fonksiyonu sayesinde rejim değişiklikleri düzenli bir biçimde meydana gelir.  $t = 1 - p, 1 - (p - 1), \dots, -1, 0, 1, \dots, T - 1, T$  zamanında gözlenen bir  $y_t$  zaman serisi için  $p$  gecikmeli STAR modeli aşağıdaki gibi tanımlanabilir (van Dijk, Teräsvirta & Franses, 2002):

$$y_t = (\phi_{1,0} + \phi_{1,1}y_{t-1} + \dots + \phi_{1,p}y_{t-p})(1 - G(s_t; \gamma, c)) + (\phi_{2,0} + \phi_{2,1}y_{t-1} + \dots + \phi_{2,p}y_{t-p})G(s_t; \gamma, c) + \varepsilon_t \quad (1)$$

veya

$$y_t = \phi'_1 x_t (1 - G(s_t; \gamma, c)) + \phi'_2 x_t G(s_t; \gamma, c) + \varepsilon_t, t = 1, \dots, T \quad (2)$$

Burada  $x'_t = (y_{t-1}, \dots, y_{t-p})'$  olmak üzere  $x_t = (1, x'_t)$ ,  $i=1,2$  için  $\phi_i = (\phi_{i,0}, \phi_{i,1}, \dots, \phi_{i,p})'$  ve  $\varepsilon_t \sim \text{nid}(0, \sigma^2)$ 'dir.  $s_t$  geçiş değişkeni,  $G(s_t; \gamma, c)$  aldığı değerler sıfır ve bir arasında değişen geçiş fonksiyonu ve  $c$  eşik değeridir.  $\gamma$  ise geçiş hızını belirleyen parametredir. Geçiş değişkeni sabit eşik değeri  $c$ 'den saptıkça düzenliliği  $\gamma$  parametresi tarafından belirlenen rejimler arası geçiş gerçekleşecektir. Bu noktada rejim değişikliği davranışı seçilen geçiş fonksiyonuna göre farklılaşacaktır. Lojistik ve üstel geçiş fonksiyonu yaygın olarak kullanılan iki geçiş fonksiyonudur. STAR modellerinde geçiş fonksiyonu olarak bunlardan ilki kullanılırsa lojistik STAR (LSTAR), ikincisi kullanılırsa üstel STAR (ESTAR) olarak adlandırılan modeller elde edilir. LSTAR ve ESTAR modelleri çok farklı dinamik ekonomik davranışı tasvir edebilirler ancak hangi modelin kullanılacağına ilişkin bir ekonomik teori bulunmamaktadır bu yüzden seçim veriye dayalı olarak yapılmalıdır (Teräsvirta & Anderson, 1992).

STAR modelleri oluşturulurken; yapısal parametreler (eşik değeri, gecikme parametresi ve  $\gamma$  parametresi) ve rejimlerin otoregresif dereceleri olmak üzere tahmin edilmesi gereken birçok serbest parametre bulunmaktadır. Dahası, bazıları kesikli (gecikme parametresi ve otoregresif dereceler), bazıları sürekli (eşik değeri ve  $\gamma$  parametresi) olan yapısal parametreleri tahmin etmek için mevcut bir analitik optimizasyon metodu yoktur (Battaglia & Protopapas, 2011a). Öte yandan, STAR modelleri yapısal parametrelerin bilinen sabit değerleri için otoregresif parametrelerde doğrusaldır, dolayısıyla en küçük kareler yöntemi ile kolaylıkla tahmin edilebilirler. Prensipte STAR modelleri, yapısal parametrelerin alabilecekleri değerlerinin tüm olası kombinasyonları dikkate alınarak bunlar içerisinde bilgi kriterine göre en iyi olanın seçilmesi yoluyla belirlenebilirler. Ancak, pratikte yapısal parametrelerin alabileceği değerlerin sayısına bağlı olarak çok fazla aday model bulunabilir ki bu durumda ortaya çıkan arama uzayının büyüklüğü tüm aday çözümlerin teker teker belirlenmesini ve makul sürelerde kontrol edilmesini olanaksız kılar (Baragona, Battaglia & Cucina, 2004). Örneğin; iki rejimli bir STAR modeli için bilinmeyen yapısal parametrelerden rejimlerin otoregresif dereceleri ( $p_1$  ve  $p_2$ ), gecikme parametresi ( $d$ ) için üst sınırı  $p_1^{maks} = p_2^{maks} = 10$  ve  $d^{maks} = 5$  şeklinde olduğu, eşik değişkeni için olası 50 değer ve  $\gamma$  parametresi için olası 100 değer bulunduğunu varsayılırsa, tam gecikme yapısında uygun modeli bulmak için toplamda 2500000 modelin denenmesi gerekecektir. Dahası eğer doğru gecikme yapısı otomatik olarak belirlenmek istenirse bu durumda 104755225000 tane STAR model denenmek zorundadır. Böyle bir durumda genel olarak değerlendirmemiz gereken  $d^{maks} \left( 2^{p_1^{maks}} - 1 \right) \left( 2^{p_2^{maks}} - 1 \right) \gamma$  tane STAR modeli vardır.

Geleneksel yöntemler ile uygun STAR modeli seçilirken, belirlenmesi gereken parametrelerden biri ya da ikisi sabitken diğerleri belirlenmekte yani ardışık bir seçim yöntemi izlenmektedir. Şöyle ki; ilk adımda, eldeki veriye uygun bir doğrusal otoregresif model Akaike bilgi kriteri (AIC), Bayesian bilgi kriteri (BIC) veya Ljung-Box istatistiği kullanılarak tahmin edilir. Bu otoregresif model için seçilen gecikme derecesi doğrultusunda, istatistiksel testler yardımıyla diğer parametreler ardışık olarak tahmin edilerek STAR modelleri belirlenir. Burada seçilen gecikme derecesinin farklı değerleri için en uygun model tamamen farklılaşacaktır. En önemlisi regresyona ilgisiz



gecikmelerin dahil edilmesi veya modele yanlış kısıtların getirilmesi, seçilen gecikme derecesi takip eden testlerin sonuçlarının belirlenmesinde esas olduğundan sürecin devamı açısından büyük ölçüde yanlış yönlendirici olacaktır (Balcombe, 2005).

Son adımda ise seçilen gecikme derecesi, geçiş değişkeni ve geçiş fonksiyonuna göre diğer parametreler bir doğrusal olmayan optimizasyon yöntemi kullanılarak tahmin edilir. Bu aşamada, eşik değeri (serinin kestiği bir değerden seçilerek) ve  $\gamma$  parametresi için kapsamlı ızgara araması gerçekleştirilerek ulaşılan parametre çifti, doğrusal olmayan optimizasyon için başlangıç değeri olarak seçilmektedir. STAR modellerinde arama uzayı çok fazla yerel optimuma sahip olma eğilimindedir (Maringer & Meyer, 2008). Bu yüzden doğrusal olmayan optimizasyon ile elde edilen katsayı tahminleri önemli ölçüde başlangıç değerlerine dolayısıyla gerçekleştirilen ızgara araştırmasının kapsamına bağlı olacaktır.

Genel olarak STAR modellerinde ilgilenilen değişkenin hangi gecikmelerinin modele dahil edileceği açık değildir. Nihai modele ulaşmak için izlenen genel yol; tam gecikme yapısı ile başlayıp daha sonra parametreleri tahmin etmeye devam ederek, tahminleri bir en düşük anlamlılık seviyesine ulaşamayan gecikmeleri çıkarmaktır. Fakat bu yaklaşımın iki ciddi eksikliği bulunmaktadır; (i) uygun gecikmeler zamanından önce dışlanabilirler ve (ii) dışlama süreci bir ya da daha fazla geçici parametre tahminleri yerel optimumu işaret ettiğinde yanlış yönlendirilmiş olabilir (Maringer & Meyer, 2008).

Yapısal parametrelerin verilen değerleri için otoregresif katsayıların tahminlerinin etkin bir şekilde gerçekleştirebilmesi nedeniyle, bir optimizasyon problemi olarak ele alındığında STAR modellerini belirleme sürecinde en önemli adım yapısal parametrelerin belirlenmesidir. Bu parametrelerin belirlenmesi süreci kesikli argümanlardan oluşan bir amaç fonksiyonunun optimum değerinin bulunmasını gerektirecektir ki (Baragona & Cucina, 2009) arama uzayındaki süreksizlikler nedeniyle bu optimizasyon problemine geleneksel sayısal metotlar uygulanamaz (Maringer & Meyer, 2008; Gilli & Winker, 2009). Eğer yapısal parametrelere bağlı olarak oluşturulan tüm aday modelleri denemek mümkün ise bu adımı geçmek çok zor

olmayacaktır. Ancak aday model sayısı arttıkça basitçe çok fazla zaman alacağı ve aşırı hesaplama maliyeti gerektireceği için tüm modelleri denemek artık uygun olmaz. Bu nedenledir ki geleneksel yöntemler kullanılırken çoğunlukla yapısal parametrelerin örneğin eşik değerinin belirlenen sayıda değeri aramaya dahil edilmekte dolayısıyla en iyi model bazı kısıtlamalar çerçevesinde belirlenmektedir. Bu yüzden yapısal parametrelerin verilen maksimum değerleri doğrultusunda, önemli gecikmeleri içeren alt modeli belirlemek için etkili bir yola ihtiyacımız vardır; yani en uygun modeli bulmak için genetik algoritmalar gibi daha spesifik bir teknik kullanmak gerekir. Genetik algoritmalar burada olduğu gibi amaç fonksiyonu düşünüm tabanlı optimizasyon metotlarını uygulamak için gerekli süreklilik, türevlenebilirlik ve içbükeylik gibi olağan matematiksel özelliklere sahip değilse bilhassa kullanışlıdır (Baragona & Cucina, 2009).

Maringer ve Meyer (2008), STAR modellerinin tahmin sürecini sezgisel algoritmalar kapsamında değerlendirmişlerdir. Özellikle STAR modellerindeki gereksiz parametreler konusunu incelemişlerdir. Çalışmaları sonucunda, hesaplama yönünden deterministik algoritmalarından daha pahalı olmayan sezgisel tekniklerin, özellikle çok fazla aday varken uygun STAR modelinin seçiminde dikkate alınması gereken bir yaklaşım olduğu görülmüştür. Battaglia ve Protopapas (2011a), çoğu zaman serisinde gözlenen doğrusal dışılık ve durağan olmama durumlarını eş zamanlı olarak ele aldıkları çalışmalarında, zaman serileri için uygun iki rejimli kendinden uyarımlı otoregresif (SETAR), STAR veya parçalı doğrusal eşik otoregresif (PLTAR) modellerinin parametrelerini model türü de dahil olmak üzere genetik algoritma kullanarak belirlemişlerdir. Önerdikleri yöntemin performansını altı simülasyon modeli ve gerçek veriler kullanarak ölçmüşler ve incelenen karmaşık modellerin genetik algoritma ile başarılı ve nispeten kolay bir şekilde belirlenebileceği sonucuna varmışlardır. Ayrıca, yazarlar söz konusu çalışmalarını STAR ve PLTAR modelleri için ikiden fazla rejimi (en fazla dört) içerecek şekilde genişletmişlerdir (Battaglia & Protopapas, 2011b). Bunların dışında rejim değişikliği modellerinin genetik algoritmalar ile ele alındığı (Wu & Chang, 2002; Baragona, Battaglia & Cucina, 2004; Baragona & Cucina, 2008), rejim değişikliği modellerinde bilgi kriterleri kullanılarak her bir rejimin otoregresif derecesi ve/veya gecikme parametresinin belirlendiği (Wong & Li, 1998; De Gooijer, 2001),

uygun rejim sayısının bulunduğu (Gonzalo & Pitarakis, 2002; Strikholm & Terasvirta, 2006; Pena & Rodriguez, 2005; Hamaker, 2009) çeşitli çalışmalarda mevcuttur.

Bu çalışmada, amaç gözlenen zaman serisine belirli kriterler dahilinde bir dizi STAR modeli içinden sadece en iyi uyan modeli seçmektir. Bu çerçevede, genetik algoritma temelli yaklaşım ile yapısal parametreler, gerekli otoregresif parametreler ve uygun STAR modelinin türü birlikte belirlenecektir. Böylece model kuranın şahsi yargılarına dayanan ayrıntılı ve zaman alıcı model belirleme süreci otomatik olarak gerçekleştirilecektir.

## 2. STAR MODELLERİNİN TAHMİNİ

Geçiş değişkeni ve geçiş fonksiyonu seçildikten sonra STAR modelinin parametreleri doğrusal olmayan en küçük kareler metodu ile tahmin edilebilir. Bu süreçte uygun herhangi bir iteratif doğrusal olmayan optimizasyon algoritması kullanılabilir. Fakat model tahmin edilirken geçiş değişkeni optimizasyonda sayısal problemlere sebep olacak kadar büyük değerlere ulaşabilir (Zivot & Wang, 2006). Söz konusu problemleri azaltmak ve yakınsamayı hızlandırmak amacıyla modelde standartlaştırılmış geçiş fonksiyonu kullanılır. Standartlaştırmadan sonra ESTAR ve LSTAR modellerinde geçiş fonksiyonları sırasıyla aşağıdaki gibidir:

$$G(s_t; \gamma, c) = 1 - \exp\{-\gamma (s_t - c)^2 / \sigma_{s_t}^2\} \quad (3)$$

$$G(s_t; \gamma, c) = (1 + \exp\{-\gamma (s_t - c) / \sigma_{s_t}\})^{-1} \quad (4)$$

Fonksiyonlarda  $\sigma_{s_t}$ , geçiş değişkeni  $s_t$ 'nin örnek standart sapmasıdır.

STAR modeli geçiş parametresi ve eşik değerinin bilinen sabit değerleri için otoregresif parametrelerde doğrusaldır (Leybourne, Newbold & Vougas, 1998). Bu yüzden  $\gamma$  ve  $c$ 'ye bağlı olarak  $\phi = (\phi_1', \phi_2)'$  en küçük kareler yöntemi ile  $x_t(\gamma, c) = (x_t'(1 - G(s_t; \gamma, c)), x_t'G(s_t; \gamma, c))'$  olmak üzere aşağıdaki gibi tahmin edilebilir (Franses & Dijk, 2000):

$$\hat{\phi}(\gamma, c) = (\sum_{t=1}^T x_t(\gamma, c)x_t(\gamma, c)')^{-1}(\sum_{t=1}^T x_t(\gamma, c)y_t) \quad (5)$$

İlgili kalıntılar  $\hat{\varepsilon}_t = y_t - \hat{\phi}(\gamma, c)'x_t(\gamma, c)$  ve kalıntı varyansı  $\hat{\sigma}^2(\gamma, c) = T^{-1} \sum_{t=1}^T \hat{\varepsilon}_t^2(\gamma, c)$  şeklinde hesaplanabilir.

STAR modelinin koşullu doğrusallığı kullanılarak takip eden  $\phi_1$  ve  $\phi_2$  parametrelerine göre kalıntı kareler fonksiyonu yazılarak tahmin problemi basitleştirilebilir (van Dijk vd., 2002):

$$Q_T(\gamma, c) = \sum_{t=1}^T (y_t - \phi(\gamma, c)'x_t(\gamma, c))^2 \quad (6)$$

Bu noktada  $Q(\gamma, c)$  fonksiyonu  $\gamma$  ve  $c$  parametrelerine göre en küçüklenecek parametre tahminlerine ulaşılır.

Doğrusal olmayan optimizasyon için uygun başlangıç değerleri geçiş parametresi ve eşik değeri için iki boyutlu ızgara araması gerçekleştirilmesi suretiyle elde edilebilir. Optimizasyon algoritmaları genel olarak parametrelerin başlangıç değerlerine çok duyarlıdır. Bu nedenle; aramada kullanılacak değerler dizisi, eşik değeri ve geçiş parametresinin seçilen her değeri için geçiş fonksiyonunun aldığı değerlerin yeterli değişkenliğe sahip olması sağlanacak şekilde belirlenmelidir (Franses & Dijk, 2000). Dolayısıyla, olası eşik değerleri dizisi geçiş değişkeninin gözlenen değerleri arasından belirlenir. Devamında kalıntı varyansının ( $\hat{\sigma}^2(\gamma, c)$ ) en küçük tahminini veren  $\gamma$  ve  $c$  çifti uygun başlangıç değerleri olarak seçilir ve doğrusal olmayan optimizasyona geçilir.

### **3. STAR MODELLERİNİN GENETİK ALGORİTMALAR İLE BELİRLENMESİ**

Davranışlarını doğadaki evrim sürecinden türeten genetik algoritmalar; arama uzaylarında süreksizliklerin söz konusu olduğu hem sürekli hem de kesikli değişkenlere sahip optimizasyon problemlerinin çözümünde, böyle problemler için etkin olmadıkları gibi hesaplama yönünden de pahalı olan geleneksel doğrusal olmayan programlama tekniklerine iyi bir alternatiftir (Bingül, Sekmen, Palaniappan & Zein-Sabatto, 2000).



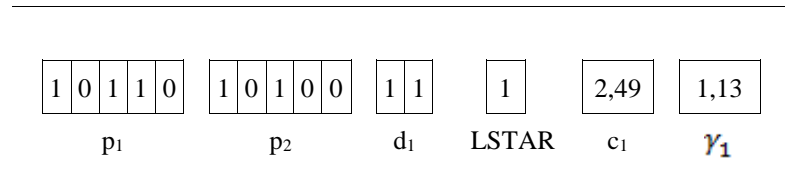
Alan bilgisi ve teorisinin sağlanmasının zor veya imkansız olduğu uygulamalar için ideal olmalarına karşın bir çok stokastik metotta olduğu gibi optimum çözümü bulmaları kesin değildir. Bu yüzden ele alınan soruna ilişkin en iyi çözümün kesinlikle ne olabileceğinin bilinmediği durumlarda kullanılması daha uygundur (Şen, 2004). Genetik algoritmalar; en iyi çözüme ulaşmak için bir popülasyondaki, her biri olası bir çözümü gösteren kromozomlara bir dizi tekrar boyunca genetik operatörleri uygular. Standart bir genetik algoritma takip eden adımlardan oluşur (Nabiyev, 2005; Sivanandam & Deepa, 2007):

- Olası çözümlerin kodlanarak ifade edildiği kromozomlardan bir popülasyon oluşturulur. Popülasyon sabit büyüklükte ya da nesilden nesle değişen büyüklükte olabilir. Büyük popülasyonlarda genetik algoritmalar global optimuma ulaşma hususunda daha etkinken artan popülasyon büyüklüğü genetik algoritmanın çalışma süresini de artırmaktadır.
- Popülasyondaki her kromozomun uygunluk değeri bulunur. Bu değerler çözülecek problemin doğasına bağlı olarak belirlenen bir uygunluk fonksiyonu kullanılarak hesaplanır. Uygunluk fonksiyonu genetik algoritmalarda probleme özel çalışan tek kısımdır ve algoritmanın başarısı çoğu zaman bu fonksiyonun verimli ve hassas olmasına bağlıdır.
- Rulet tekerleği seçilimi veya turnuva seçilimi gibi yöntemlerle seçilen kromozomlar eşlenerek, çaprazlama ve mutasyon operatörlerinin uygulanması suretiyle yeni popülasyon oluşturulur.
- Popülasyonda yeni kromozomlar eski kromozomlar ile yer değiştirir ve bu sayede sabit büyüklükte bir popülasyon sağlanır.
- Yeni popülasyondaki kromozomların uygunluk değerleri hesaplanarak yeni popülasyonun başarısı bulunur.
- Bu noktaya kadar olan işlemler belirli sayıda tekrarlanarak daha iyi nesiller elde edilir.
- Çözüm popülasyonlar hesaplanırken bulunan en iyi kromozomdur.

STAR modelleri genetik algoritmalar yardımıyla belirlenirken rejim sayısının bilindiği varsayılacaktır. STAR modellerinde kromozom formunda temsil edilmesi



gereken deęişkenler; eşik deęeri, gecikme parametresi,  $\gamma$  parametresi, rejimlerin otoregresif dereceleri ve model türüdür. LSTAR ve ESTAR olmak üzere dikkate alınan iki tür model olduęu için kromozomda model türü bir bit yardımıyla kodlanmıştır. Söz konusu bit LSTAR modeli için bir deęerini alırken ESTAR modeli için sıfır deęerini almaktadır. Gerek  $\gamma$  gerekse eşik deęeri sürekli deęişkenler olmaları itibariyle, her iki parametre için en doęal kodlama deęer kodlamasıdır. Dolayısıyla bu parametreler hiç bir dönüşüme tabii tutulmadan doğrudan reel sayı olarak kromozomlarda yer almışlardır. Diğer taraftan bağımlı deęişkenin gecikmeli deęerlerinin hangisinin modelde yer alıp almayacağı iki seçenekli bir durumdur; dolayısıyla otoregresif parametreler için en doęal kodlama ikili kodlamadır. Her bir rejim için ilk bit sabit terimi gösterirken, ikinci bitten itibaren  $(i+1)$ . bit o rejimdeki  $i$ . otoregresif parametreyi temsil etmektedir. Bitlerin bir deęeri alması deęişkenin modelde yer aldığı, sıfır deęeri alması modelde yer almadığı anlamına gelmektedir. Gecikme parametresi  $d$  ise, tam sayı olması itibariyle ondalık sayı sisteminden ikilik sayı sistemine çevrilerek kodlanmıştır. Açıklamalar çerçevesinde iki rejimli STAR modelleri için örnek bir kodlama Şekil 1'de gösterilmiştir.



**Şekil 1: İki Rejimli STAR Modeli İçin Kromozom Kodlaması**

#### 4. UYGULAMA

Açıklanan kodlama stratejisi kullanılarak genetik algoritmanın STAR modellerini belirleme noktasındaki başarısı, 1968:06–1989:12 dönemini kapsayan aylık ABD işsizlik oranları verileri kullanılarak sınanmıştır. Bu verileri kullanarak van Dijk vd. (2002) geleneksel modelleme yöntemi için örnek bir analiz gerçekleştirmişlerdir.

Yazarların önerdikleri LSTAR spesifikasyonu doğrultusunda yayınladıkları program kodları kullanılarak takip eden model tahmin edilmiştir.<sup>1</sup>

$$\begin{aligned}
 \Delta y_t = & 0,479 + 0,645D_{1,t} - 0,342D_{2,t} - 0,680D_{3,t} - 0,725D_{4,t} - 0,649D_{5,t} \\
 & (0,064) (0,065) (0,097) (0,081) (0,102) (0,090) \\
 & -0,317D_{6,t} - 0,410D_{7,t} - 0,501D_{8,t} - 0,554D_{9,t} - 0,306D_{10,t} \\
 & (0,081) (0,086) (0,079) (0,086) (0,066) \\
 & +[-0,040y_{t-1} - 0,146\Delta y_{t-1} - 0,101\Delta y_{t-6} + 0,097\Delta y_{t-8} \\
 & (0,009) (0,078) (0,060) (0,056) \\
 & -0,123\Delta y_{t-10} - 0,129\Delta y_{t-13} - 0,103\Delta y_{t-15}][1 - G(\Delta_{12}y_{t-1}; \gamma, c)] \\
 & (0,056) (0,063) (0,059) \\
 & +[-0,011y_{t-1} + 0,225\Delta y_{t-1} + 0,307\Delta y_{t-2} - 0,119\Delta y_{t-7} \\
 & (0,008) (0,078) (0,076) (0,067) \\
 & -0,155\Delta y_{t-13} - 0,215\Delta y_{t-14} - 0,235\Delta y_{t-15}]G(\Delta_{12}y_{t-1}; \gamma, c) + \varepsilon_t \\
 & (0,089) (0,085) (0,085)
 \end{aligned} \tag{7}$$

$$G(\Delta_{12}y_{t-1}; \gamma, c) = \left(1 + \exp\left\{-23.155(\Delta_{12}y_{t-1} - 0,273)/\sigma_{\Delta_{12}y_{t-1}}\right\}\right)^{-1} \tag{8}$$

(20.682) (0.054)

$$\sigma = 0,196, AIC = -3,154, BIC = -2,763$$

Eşitlikte  $D_{i,t}$ ,  $i = 1, \dots, 11$  olmak üzere t. gözlem i. aya ait ise bir diğer durumda sıfır değerini alan mevsimsel kukla değişkendir.

(7), genetik algoritmanın performansını değerlendirme hususunda referans olarak kullanılacaktır. Mevcut veri seti için uygun STAR modeli genetik algoritma yardımıyla tahmin edilecek ve bu model verilen referans model ile karşılaştırılarak algoritmanın etkinliği araştırılacaktır. STAR tipi doğrusal dışılığın varlığının teyit edildiği varsayımı altında referans modele ulaşma süreci üç farklı açıdan ele alınacaktır. İlk önce referans modelin gecikme yapısı, kukla değişkenleri, gecikme parametresi ve model türü olduğu gibi korunarak genetik algoritma bir doğrusal olmayan optimizasyon algoritması olarak değerlendirilip eşik değeri ve  $\gamma$  parametresi tahmin edilecektir

<sup>1</sup> Yazarların çalışmalarında raporladıkları değerler ile dağıttıkları program kodları ile ulaşılan değerler birbirinden farklı olduğu için çalışmada programdan elde edilenler kullanılmıştır.

(Deney 1). Çünkü bilindiği üzere tahmin için yaygın olarak kullanılan bu tür algoritmaların sonuçları başlangıç değerlerinin seçiminden önemli ölçüde etkilenmektedirler (Maringer & Meyer, 2008). Devamında gecikme parametresi, rejimlerin gecikme yapısı ve kukla değişkenler (Deney 2), son olarak da model türü (Deney 3) aramaya dahil edilecektir. Birinci deneyde kalıntı kareler toplamı (KKT) en küçüklenecek sonraki deneylerde ise  $\hat{\sigma}$  ile AIC ve BIC kriterlerine göre ayrı ayrı en uygun modeller belirlenecektir. KKT (6) kullanılarak,  $\hat{\sigma}$  ise k modeldeki parametre sayısı ve  $\hat{\varepsilon}_t$  (5)'den elde edilen kalıntılar olmak üzere aşağıdaki gibi hesaplanmıştır:

$$\hat{\sigma} = \sqrt{\frac{1}{T-k} \sum_{t=1}^T \hat{\varepsilon}_t^2} \quad (9)$$

AIC ve BIC hesaplanırken aşağıda verilen genel formül kullanılmıştır (Balcombe, 2005):

$$C(k) = \ln \hat{\sigma}^2 + f(k, T) \quad (10)$$

Eşitlikte,  $f(k, T)$  gözlem ve tahmin edilen parametre sayısına göre belirlenen ceza faktörüdür. Bu faktör AIC kriteri için  $f(k, T) = 2k/T$  ve BIC için  $f(k, T) = \ln(T)k/T$ 'dir.

Uygunluk değeri olarak kullanılmadan önce bir s kromozomu ile belirlenen x modelinden hesaplanan  $\hat{\sigma}$ , AIC ve BIC takip eden şekilde dönüştürülmüştür:

$$f(s) = \begin{cases} C_{maks} - C(x), & \text{eğer } C(x) < C_{maks} \\ 0, & \text{d. d.} \end{cases} \quad (11)$$

Burada  $C_{maks}$  ilgili kriterinin mevcut popülasyona kadar ki en büyük değeridir. Devamında dönüştürülen uygunluk değeri,  $f(s)^{1,005}$  şeklinde ölçeklendirilmiştir.

İlk deneyde eşik değeri ve  $\gamma$  parametresi için değer kodlaması kullanılmıştır. Parametrelerin alt ve üst sınırları,  $\gamma$  için  $1 \leq \gamma \leq 1000$  şeklinde, eşik değeri için ise



gözlenen geçiş değişkeninin ilk ve son %10'luk kısımları dışlanarak belirlenmiştir. 100 kez çalıştırılan genetik algorithmadan elde edilen sonuçlar Tablo 1'de verilmiştir.

**Tablo 1: Birinci Deneye İlişkin Sonuçlar**

	KKT	AIC	BIC	$\gamma$	c	
Referans	8,178316	-3,154153	-2,762581	23,154887	0,273692	
GA	En iyi	8,178316	-3.154153	-2.762581	23,154895	0,273692
	En kötü	8,178316	-3.154153	-2.762581	23,154824	0,273693

Algoritma tüm denemelerde virgülden sonra onikinci haneye kadar birbirlerine eşit KKT değerleri hesaplamıştır ki bu değerde aynı hassasiyet düzeyinde referans modelden elde edilen KKT değerine eşittir. Aynı şekilde  $\gamma$  ve c için sırasıyla virgülden sonra dördüncü ve beşinci haneye kadar eşit değerler hesaplanmıştır. Tahmin edilen parametreler virgülden sonra dört basamak dikkate alındığında referans değerlerine eşittirler. Zaten bu parametrelerin yüksek hassasiyette tahmin edilmesine gerek yoktur, çünkü bunlardan birinde gerçekleşecek büyük bir değişimin geçiş fonksiyonu üzerindeki etkisi çok küçük olacaktır (van Dijk vd., 2002). Dolayısıyla genetik algoritma, otoregresif parametreleri tahmin etmek amacıyla kullanılan doğrusal olmayan optimizasyon algoritmaları için anlamlı başlangıç değerleri bulmanın ötesinde söz konusu parametrelerin nihai tahminlerine ulaşmak içinde etkin bir şekilde kullanılabilir.

İkinci ve üçüncü deneyde, gen büyüklüklerinin ele alınan probleme göre yeniden belirlendiği Şekil 1'de verilen kromozom kodlaması kullanılmıştır. Yalnız ikinci deneyde model türünün belirlenmesi amacıyla kullanılan gen kromozoma dahil edilmemiştir. Bu şekilde kodlanan kromozomlar ile  $\gamma$  ve c belirli bir hassasiyette tahmin edilirken, diğer parametrelerin tahminlerine kesin olarak ulaşılabilir. Çünkü  $\gamma$  ve c dışındaki yapısal parametrelerin uygunluk fonksiyonu üzerinde etkisi çok büyüktür. Bunlardan birinde meydana gelecek bir değişim örneğin gecikme yapısına bir değişkeni eklemek ya da çıkarmak tüm mevcut parametre değerlerini bir anda kullanışsız hale getirebilir (Maringer & Meyer, 2008). Bu yüzden ikinci ve üçüncü deneylerde, ilk genetik algoritma işletildikten sonra diğer tüm parametreler ilk algorithmadan elde edilen

sonuçlara göre sabitlenerek birinci deneydeki tasarıma ve kodlamaya sahip ikinci bir genetik algoritma ile  $\gamma$  ve c parametrelerinin nihai tahminleri elde edilmiştir. Eşik değerinin sınırları ilk deneydeki gibi belirlenirken,  $\gamma$  için alt ve üst sınırlar gerçekleştirilen bir ön çalışma doğrultusunda ilk genetik algortmada  $1 \leq \gamma \leq 500$ , ikincisinde ise  $1 \leq \gamma \leq 4000$  olarak seçilmiştir. Van Dijk vd. (2002)'de olduğu gibi maksimum gecikme uzunluğu otoregresif parametreler için 15, gecikme parametresi için ise 6 olarak alınmıştır. Tüm modeller 1970:01 ve 1989:12 dönemleri arasını kapsayan 240 efektif gözlem kullanılarak tahmin edilmiştir. Arama uzayının çok geniş olması ve genetik algoritmaların diğer sezgiseller gibi yerel optimum değerlere yakınsayabilmesi nedeniyle her deney için tekrarlı denemeler yapılmıştır. Bu doğrultuda, mevcut veri setinde algoritma her amaç fonksiyonu için on kez çalıştırılmış ve ulaşılan en iyi modeller Tablo 2'de verilmiştir.

Tablo 2 incelendiğinde hem ikinci hem de üçüncü deneyde genetik algoritmanın, kullanılan üç farklı amaç fonksiyonu açısından da farklı modelleri en iyi model olarak belirlediği görülmektedir. Öncelikle en iyi LSTAR modelinin arandığı ikinci deneyde bütün modeller referans modelle aynı mevsimsel kukla değişkenleri içermektedir. BIC kriteri açısından referans modelle aynı gecikme parametresine sahip olsa da gecikme yapısı bakımından tamamen farklı bir model seçilmiştir. Seçilen modelin doğrusal bölümü hiç bir gecikme içermemektedir.  $\phi$  ve AIC kriteri dikkate alındığında, doğrusal kısımda AIC kriterinin beşinci gecikmeyi içermemesi dışında birbirine benzer, referans modelden tamamen farklı, daha cimri modeller seçilmiştir. Üçüncü deneyle beraber aramaya model türünün dahil edilmesi ulaşılan nihai modellerin uyumunu önemli ölçüde artmıştır. Ancak tahmin edilen modellerin tümü yapısal parametreler ve mevsimsel kukla değişkenler noktasında referans modelden tamamen ayrılmaktadır. Bu deneyde tüm kriterler veri setine ESTAR türünde bir modelinin daha iyi uydurulabileceğini işaret etmektedir ki bu şekilde tahmin edilen modellerden hesaplanan  $\phi$ , AIC ve BIC değerleri, ikinci deneyde söz konusu kriterlerin her biri için elde edilen en iyi modellerden hesaplanana göre daha düşüktür. İkinci deneyde olduğu gibi bu deneyde de  $\phi$  ve AIC kriteri kullanılarak tahmin edilen modeller birbirine benzemektedir, ancak bu kez tahmin edilen modelin sahip olduğu parametre sayısına bağlı olarak cezalandırılmamasının bir sonucu olarak,  $\phi$  ile belirlenen modelde



AIC kriterine kullanılarak belirlenene göre daha fazla gecikme yer almıştır. BIC kriteri kullanılarak ise bu modellerden farklı gecikme parametresi olan ve daha az gecikme içeren bir modele ulaşılmıştır.

Farklı amaç fonksiyonları için farklı modellerin belirlenmesi durumu ilk bakışta bir dezavantaj gibi görünmektedir. Ancak her iki deneyde de genetik algoritma temelli yaklaşım ile tüm amaç fonksiyonları için referans olarak alınan modelden daha iyi uyuma sahip modellere ulaşıldığı dikkate alındığında, bu durumun modeli belirleme amacı doğrultusunda bir tercihe olanak sağlaması açısından bir avantaj olarak düşünülmesi daha doğrudur.

Genel olarak ilk aşamadaki genetik algoritmanın ulaştığı modeller referans modellerden daha iyi olsa da mevcut modeller, ikinci bir genetik algoritma kullanılarak kayda değer oranda iyileştirilebilmiştir. Dolayısıyla daha önce bahsedilen nedenlerden dolayı STAR modellerinin tüm yapısal parametrelerinin genetik algoritma ile eş zamanlı belirlenmesi noktasında önceliğin  $\gamma$  ve eşik değeri dışındaki parametrelere verilmesi, bu iki parametrenin diğer parametrelerin değerlerine göre en iyilenmesi mantıklı ve geçerli bir yaklaşımdır.

Şüphesiz genelden özele yaklaşımı yerine, gerçek gecikme yapısını doğrudan diğer yapısal parametreler ile birlikte belirlemeye yönelik bir yaklaşımın kullanılması elde edilen modellerin kalitesini arttırmaktadır. Ancak belirlenen bu modellerin, herhangi bir kriter açısından en iyi uyum sağlayan modeller olduğundan tüm olası modelleri tahmin etmeden emin olamayız. Nitekim her bir amaç fonksiyonu için gerçekleştirilen on tekrarın tamamında algoritma, aralarında genellikle küçük farklılıklar da olsa değişik modellere yakınsamıştır. Ayrıca elde edilen modellerden bazıları referans modelin sahip olduğu uyuma dahi ulaşamamıştır. Ancak bu genetik algoritma temelli model belirleme yaklaşımının geleneksel deterministik yaklaşımlara, bir kaç tekrardan sonra en az onların ulaştığı kadar iyi bir modeli belirleyebilmesi noktasında bir alternatif olarak düşünülmesine engel değildir.

**Tablo 2: İkinci ve Üçüncü Deney Sonucunda Tahmin Edilen En İyi Modeller**

Deney	Amaç Fonk.	Model	$\phi_{1,i}$	$\phi_{2,i}$	$D_i$	d	c	$\gamma$	Model Türü	$\delta$	KKT	AIC	BIC
		Ref.	1-6-8-10-13-15	1-2-7-13-14-15	1-2-3-4-5-6-7-8-9-10	1	0,273692	23,154887	LSTAR	0,195949	8,178316	-3,154153	-2.762581
	$\delta$	GA.1	5-8	1-2-3-5-7-9-13-14-15	1-2-3-4-5-6-7-8-9-10	2	0,253135	443,771839	LSTAR	0,194789	8,119785	-3,169667	-2,792599
		GA.2	5-8	1-2-3-5-7-9-13-14-15	1-2-3-4-5-6-7-8-9-10	2	0,254674	2986,388556	LSTAR	0,194784	8,119321	-3,169726	-2,792657
2	AIC	GA.1	8	1-2-3-5-7-9-13-14-15	1-2-3-4-5-6-7-8-9-10	2	0,253754	369,847836	LSTAR	0,194928	8,169321	-3,171920	-2,809353
		GA.2	8	1-2-3-5-7-9-13-14-15	1-2-3-4-5-6-7-8-9-10	2	0,248667	1411,905903	LSTAR	0,194921	8,168721	-3,171993	-2,809427
	BIC	GA.1	-	2-7-14-15	1-2-3-4-5-6-7-8-9-10	1	0,616222	239,490805	LSTAR	0,200946	8,923790	-3,133585	-2,858034
		GA.2	-	2-7-14-15	1-2-3-4-5-6-7-8-9-10	1	0,623647	3249,796840	LSTAR	0,200822	8,912781	-3,134819	-2,859269
	$\delta$	GA.1	1-2-5-7-9-11-15	2-3-6-8-11-15	1-2-3-4-5-7-8-9-10-11	5	0,475869	80,054711	ESTAR	0,193235	7,916010	-3,178418	-2,772344
		GA.2	1-2-5-7-9-11-15	2-3-6-8-11-15	1-2-3-4-5-7-8-9-10-11	5	0,476583	76,183748	ESTAR	0,193219	7,914743	-3,178578	-2,772504
3	AIC	GA.1	1-4-7-13	2-3-11-15	1-2-3-4-5-7-8-9-10-11	5	0,486491	109,176290	ESTAR	0,194283	8,190844	-3,185955	-2,852394
		GA.2	1-4-7-13	2-3-11-15	1-2-3-4-5-7-8-9-10-11	5	0,477041	77,024985	ESTAR	0,193418	8,118049	-3,194882	-2,861321
	BIC	GA.1	3-6-15	2-3-9-11-14	1-3-4-5-9-11	6	0,345520	240,985174	ESTAR	0,196425	8,526777	-3,179094	-2,903544
		GA.2	3-6-15	2-3-9-11-14	1-3-4-5-9-11	6	0,345024	258,384043	ESTAR	0,196400	8,524615	-3,179348	-2,903797

Son olarak her bir kriter için ulaşılan en iyi modellere ilişkin sonuçları tamamlaması açısından kalıntılara dayalı bazı temel istatistikler Tablo 3'de verilmiştir. Tahmin edilen modele ilişkin söz konusu istatistiklere göre %5 anlam düzeyinde modelde otokorelasyon ve değişen varyans sorunları yoktur. Ayrıca tüm modeller için normallik testinin sonuçları referans modelinkinden farklı değildir. Dolayısıyla elde edilen modellerin bu istatistikler çerçevesinde referans modelin yerine tercih edilmemesi için hiçbir neden bulunmamaktadır.

**Tablo 3: En İyi Modellere İlişkin Yanlış Belirleme Testleri**

	Deney 2				Deney 3		
	Ref	AIC	BIC	$\hat{\sigma}$	AIC	BIC	$\hat{\sigma}$
Lomnicki-Jarque-Bera	19,4612 (0,0000)	15,6047 (0,0004)	35,5462 (0,0000)	14,0143 (0,0009)	12.2053 (0,0022)	6.7575 (0,0340)	13.8734 (0,0009)
Çarpıklık	0,6379	0,5346	0,6620	0,5026	0,5372	0,4067	0,5592
Basıklık	3,5643	3,6459	4,3422	3,6250	3,2569	3,1181	3,3693
LM(2)	1,0490 (0,3521)	0,5909 (0,5546)	0,0919 (0,9121)	0,6419 (0,5273)	0,3288 (0,7201)	0,3031 (0,7387)	0,1370 (0,8720)
LM(12)	0,4733 (0,9284)	0,3464 (0,9791)	0,4581 (0,9366)	0,3979 (0,9631)	0,8323 (0,6169)	0,9647 (0,4839)	0,5833 (0,8540)
ARCH(1)	0,8628 (0,3529)	1,4762 (0,2243)	1,1811 (0,2771)	1,8011 (0,1795)	1,4995 (0,2207)	0,4876 (0,4849)	1,6318 (0,2014)
ARCH(4)	1,3192 (0,8581)	3,9042 (0,4191)	1,8721 (0,7592)	4,0525 (0,3989)	8,2108 (0,0841)	3,8542 (0,4260)	6,3648 (0,1735)

Not. p değerleri parantez içerisinde verilmiştir.

## 5. SONUÇ

Önerilen genetik algoritma temelli yaklaşım ile STAR modellerini belirleme süreci otomatik olarak gerçekleştirilmiştir. Bu yaklaşımın en önemli iki özelliği; geleneksel yöntemlerde olduğu gibi herhangi bir parametrenin bir değere sabitlenmesine gerek duyulmaması, bu sayede en iyi model aranırken tüm parametrelerin içerdiği belirsizliklerin dikkate alınabilmesi ve olası model sayısı çok fazla olsa da başarılı bir şekilde uygulanabilmesidir. Ayrıca sunulan yaklaşımın nihai modelin yuvalandığı genel model yerine doğrudan rejimlerin gerçek gecikme yapılarını belirlemeye yönelik bir arama gerçekleştirmesi, bir diğer önemli özelliğidir. Diğer taraftan model belirleme süreci için



önerilen iki aşamalı genetik algoritma tasarımı ve kullanılan melez kromozom gösterimi genetik algoritma bağlamında önerilen yaklaşımın diğer önemli özellikleridir.

Çalışmada, literatürden seçilen bir referans modele ulaşma süreci üç farklı açıdan ele alınmıştır. Buna göre ilk deneyde genetik algoritmanın otoregresif parametreleri tahmin etmek amacıyla etkin bir şekilde kullanılabilceği sonucuna varılmıştır. Sonraki iki deneyin her ikisinde de seçilen kriterler açısından referans modelden daha iyi modellere ulaşılmıştır. Bu deneylerde genetik algoritma temelli bir yaklaşımda yapısal parametrelerden önceliğin  $\gamma$  ve eşik değeri dışındakilere verilerek sürecin iki aşamalı bir genetik algoritma tasarımı ile değerlendirilmesinin daha uygun olacağı düşünülmüştür. Gerçekte de bu düşüncenin ulaşılan nihai modellerin uyumu açısından tek genetik algoritmanın işletildiği sürece göre kayda değer bir iyileşmeye neden olduğu görülmüştür.

## KAYNAKÇA

Balcombe, K. G. (2005). Model Selection Using Information Criteria and Genetic Algorithms. *Computational Economics*, 25(3), 207-228.

Baragona, R., Battaglia, F., & Cucina, D. (2004). Fitting piecewise linear threshold autoregressive models by means of genetic algorithms. *Computational Statistics & Data Analysis*, 47(2), 277-295.

Baragona, R., & Cucina D. (2009). Genetic search for threshold parameters in time series threshold models: algorithms and computer programs, TECHNICAL REPORT 10/2009, *Department of Statistics, Probability and Applied Statistics, Sapienza University of Rome, Italy*. Retrieved from <http://w3.uniroma1.it/statstsmeh/download/gasdtgarch05.pdf> (12.07.2011)

Baragona, R., & Cucina, D. (2008). Double threshold autoregressive conditionally heteroscedastic model building by genetic algorithms. *Journal of Statistical Computation and Simulation*, 78(6), 541-558.



Battaglia, F., & Protopapas, M. K. (2011a). Time-varying multi-regime models fitting by genetic algorithms. *Journal of Time Series Analysis*, 32(3), 237-252.

Battaglia, F., & Protopapas, M. K. (2011b). Multi-regime models for nonlinear nonstationary time series. *Computational Statistics*.

Bingül, Z., Sekmen, A. S., Palaniappan, S., & Zein-Sabatto, S. (2000). Genetic algorithms applied to real time multiobjective optimization problems. Southeastcon 2000. Proceedings of the IEEE (pp. 95-103). Presented at the Southeastcon 2000. Proceedings of the IEEE.

Chan, K. S., & Tong, H. (1986). On Estimating Thresholds In Autoregressive Models. *Journal of Time Series Analysis*, 7(3), 179-190.

Chen, Y. T. (2003). Discriminating between competing STAR models. *Economics Letters*, 79(2), 161-167.

De Gooijer, J. G. (2001). Cross-validation Criteria for Setar Model Selection. *Journal of Time Series Analysis*, 22(3), 267-281. doi:10.1111/1467-9892.00223

Franses, P. H., & Dijk, D. van. (2000). *Nonlinear time series models in empirical finance*. Cambridge University Press.

Gilli M. & Winker P. (2009). Heuristic optimization methods in econometrics. In: Belsley, D. A., & Kontoghiorghes, E. J. (ed). *Handbook of Computational Econometrics*. John Wiley and Sons, 81-119.

Gonzalo, J., & Pitarakis, J.-Y. (2002). Estimation and model selection based inference in single and multiple threshold models. *Journal of Econometrics*, 110(2), 319-352.

Hamaker, E. L. (2009). Using information criteria to determine the number of regimes in threshold autoregressive models. *Journal of Mathematical Psychology*, 53(6), 518-529.





Wong, C. S., & Li, W. K. (1998). A note on the corrected Akaike information criterion for threshold autoregressive models. *Journal of Time Series Analysis*, 19(1), 113–124. doi:10.1111/1467-9892.00080

Wu, B., & Chang, C.-L. (2002). Using genetic algorithms to parameters (d,r) estimation for threshold autoregressive models. *Computational Statistics & Data Analysis*, 38(3), 315-330.

Zivot, E., & Wang, J. (2006). *Modeling financial time series with S-PLUS*. Birkhäuser.